

REMARKS

The substitute specification enclosed herein contains no new matter.

Claims 1, 2, 5, 8, 10-18, 20, 23, 25, 27, and 30 remain in the application and have been amended. Claims 33-56 are new. Claims 3, 4, 6, 7, 9, 19, 21, 22, 24, 26, 28, 29, 31, and 32 have been canceled.

Applicants submit herewith a substitute specification along with a redline version showing minor revisions to correct typographical and grammatical errors. No new matter has been added. Approval and entry of the substitute specification is respectfully requested.

In the Office Action mailed May 7, 2004, the Examiner objected to the Abstract as not being in single paragraph format. Applicants have amended the Abstract to now be in single paragraph format and to be within the range of 50-150 words.

Claims 2 and 4 were objected to because of informalities. Claim 2 has been amended, and claim 4 has been canceled.

Claims 1-32 were rejected under 35 U.S.C. § 102(e) as anticipated by U.S. Patent No. 6,236,395 ("Sezan et al.").

Applicants respectfully disagree with the basis for the rejection and request reconsideration and further examination of the claims.

Sezan et al. describe general ideas for providing description schemes in terms of user, system, and program. Sezan et al. describe the possibility that each description scheme or a combination thereof in the three aspects can make the user search, browse, and filter programs in a personalized manner. As described therein, Sezan et al. disclose a highlight view DS and Key frame DS together with a thumbnail view, event view, close-up view, and an alternative view DS under a visualization view DS, which is illustrated in Figure 14. Such various DS views are all independent of each other because each DS view is made for its own viewing purpose. In other words, each DS view lacks the intrinsic interrelationship among the other DS views. The scheme taught by Sezan et al. has drawbacks, including, for example, that the amount of metadata to be generated, held, and parsed, as well as the parsing time for parsing this data will increase, which causes inefficient use of resources and ultimately impacts available resources at the environmental level.

Claim 1 of the present application is directed to a HierarchicalSummary description scheme (DS) for describing a video summary wherein the HierarchicalSummary DS includes a HighlightLevel DS that includes at least one HighlightSegment DS. Moreover, in claim 1 the highlight segment DS includes information on a highlight segment corresponding to one of a plurality of video summary intervals. In particular, the HighlightSegment DS includes a VideoSegmentLocator DS describing time information or a video itself of the HighlightSegment and an ImageLocator DS describing a representative frame of the highlight segment. The HighlightSegment DS includes a video segment locator DS and an image locator DS that establishes the intrinsic relationship between a video segment locator DS and an image locator DS. This enables one DS to reference the other DS dynamically without requiring any additional efforts in terms of the system or the user.

Thus, the highlight segment in claim 1 is an optimized and especially devised description scheme that gives many possibilities. For example, multiple key frames may be associated with a single highlight segment, and thus multiple access points to a single highlight segment can be given for flexible dynamic browsing. When amending or updating the DS (for example adding key frame in a certain segment), the claimed DS does not give any difficulty to the DS generating tool.

In view of the foregoing, applicants respectfully submit that claim 1 and all claims depending therefrom are clearly in condition for allowance.

Independent claims 17, 18, 25, 44, 47, 51, and 54 all include the feature argued above with respect to claim 1. Applicants respectfully submit that these independent claims and all claims depending therefrom are allowable over the Sezan et al. reference for the reasons discussed above with respect to claim 1. In conclusion, the highlight segment DS of the present invention is characterized in that it includes a video segment locator DS describing time information or a video itself of the highlight segment and an image locator DS describing a representative frame of the highlight segment. Sezan et al. fail to disclose or suggest such features and the many advantages thereof of the present invention.

Applicants respectfully submit all of the claims in this application are now in condition for allowance. In the event the Examiner finds minor informalities that can be

resolved by telephone conference, the Examiner is urged to contact applicants' undersigned representative by telephone at (206) 622-4900 in order to expeditiously resolve prosecution of this application. Consequently, early and favorable action allowing these claims and passing this case to issuance is respectfully solicited.

The Director is authorized to charge any additional fees due by way of this Amendment, or credit any overpayment, to our Deposit Account No. 19-1090.

Respectfully submitted,

SEED Intellectual Property Law Group PLLC



E. Russell Tarleton

Registration No. 31,800

ERT:alb

Enclosure:

Postcard

Redlined Substitute Specification

Substitute Specification

701 Fifth Avenue, Suite 6300
Seattle, Washington 98104-7092
Phone: (206) 622-4900
Fax: (206) 682-6031

300055.468 / 529878_1.DOC

VIDEO SUMMARY DESCRIPTION SCHEME AND METHOD AND SYSTEM OF VIDEO SUMMARY DESCRIPTION DATA GENERATION FOR EFFICIENT OVERVIEW AND BROWSING

TECHNICAL FIELD

5 The present invention relates to a video summary description scheme for efficient video overview and browsing, and also relates to a method and system of video summary description generation to describe video summary according to the video summary description scheme.

10 The technical fields in which the present invention is involved are ~~content based-content-based~~ video indexing and browsing/searching and summarizing video to the content ~~based~~ and then describing it.

BACKGROUND OF THE INVENTION

15 The format of summarizing video largely falls into dynamic summary and static summary. The video description scheme according to the embodiments of the present invention is for efficiently describing the dynamic summary and the static summary into in the ~~unification-based-unification-based~~ description scheme.

20 Generally, because the existing video summary and description scheme provide simply the information of video interval which is included in the video summary, the existing video summary and description scheme are limited to conveying overall video contents through the playing of the ~~summary-video-video summary~~.

 However, in many cases, the browsing for identifying and revisiting concerned parts through overview of overall contents is needed rather than only overview of overall contents through the ~~summary-video-video summary~~.

25 Also, the existing video summary provides only the video interval which is considered to be important according to the criteria determined by the video summary provider. Accordingly, if the criteria of users and the video provider are different from

each other or users have special criteria, the users ~~can not~~ cannot obtain the video summary of their desires they desire.

That is, although the existing video summary video permits the users selecting the video summary video with a desired level by providing several levels' summary videos, video summary, it makes the selecting extent of the users ~~to be limited so~~ that the users ~~can not~~ cannot select by the contents of the summary videos, video summary.

The US patent 5,821,945 entitled "Method and apparatus for video browsing based on content and structure" represents video in compact form and provides browsing functionality accessing to the video with desired content through the representation.

However, the patent ~~is on the~~ pertains to static summary based on the representative frame, and although the existing static summary summarizes by using the representative frame of the video shot, the representative frame of this patent provides only visual information representing the shot, ~~the~~ The patent has a limitation on conveying the information using the summary scheme.

As compared with the patent, the video description scheme and browsing method of the embodiments described herein utilize the dynamic summary based on the video segment.

The video summary description scheme was proposed by the MPEG-7 Description Scheme (V0.5) announced ISO/IEC JTC1/SC29/WG11 MPEG-7 Output Document No. N2844 on July 1999. Because the scheme describes the interval information of each video segment of dynamic summary video, video summary, in spite of providing basic functionalities describing dynamic summary, the scheme has ~~problem~~ problems in the following aspects.

First, there is the drawback that it ~~can not~~ cannot provide access to the original video from summary segments constituting the summary video, video summary. That is, ~~the when~~ users ~~wanted~~ want to access ~~to the~~ original video to understand more detailed information on the basis of the summary contents and overview through summary video, video summary, ~~however the existing scheme could not~~ cannot meet the need.

Secondly, the existing scheme ~~can not~~ cannot provide sufficient audio summary description functionalities.

And finally, there is the drawback that in the case of representing ~~event based event-based~~ summary, the duplicate description and the complexity of searching is indispensable.

SUMMARY OF THE INVENTION

~~An object of the~~ The disclosed embodiments of the present invention is to provide a hierarchical video summary description scheme, which comprises the representative frame information and the representative sound information at each video interval ~~which that~~ is included in the video summary ~~video~~ and makes feasible the user customized ~~user-customized event-based event-based~~ summary providing the users' selection for the contents of the video summary ~~video~~ and efficient browsing ~~to be feasible~~, and a video summary description data generation method and system using the description scheme.

In order to achieve the ~~object~~ foregoing, the HierarchicalSummary DS according to an executable example of the present invention comprises at least one HighlightLevel DS, which is describing highlight level, and the HighlightLevel DS comprises at least a HighlightSegment DS, which is describing highlight segment information constituting the video summary ~~video~~ of the highlight level.

Preferably, the HighlightLevel DS is composed of at least one lower level HighlightLevel DS's.

More preferably, the HighlightSegment DS comprises a VideoSegmentLocator DS, which is describing time information or video itself of ~~said the~~ corresponding highlight segment.

It is preferable that the HighlightSegment DS further comprises ImageLocator DS, which is describing the representative frame of ~~said the~~ corresponding highlight segment.

It is more preferable that the HighlightSegment DS further comprises SoundLocator DS, which is describing the representative sound information of said corresponding highlight segment.

- 5 Preferably, the HighlightSegment DS further comprises ImageLocator DS, which is describing the representative frame of ~~said~~the corresponding highlight segment, and SoundLocator DS, which is describing the representative sound information of ~~said~~the corresponding highlight segment.

- 10 More preferably, the ImageLocator DS describes time information or image data of the representative frame of video interval corresponding to ~~said~~the corresponding highlight segment.

Preferably, the HighlightSegment DS further comprises AudioSegmentLocator DS, which is describing the audio segment information constituting an audio summary of ~~said~~the corresponding highlight segment.

- 15 More preferably, the AudioSegmentLocator DS describes time information or audio data of the audio interval of ~~said~~the corresponding highlight segment.

It is preferable that the HierarchicalSummary DS ~~includes—include~~ SummaryComponentList describing and enumerating all of the SummaryComponentTypes ~~which—that~~ is included in the HierarchicalSummary DS.

- 20 Also, it is preferable that the HierarchicalSummary DS ~~includes—include~~ SummaryThemeList DS, which is enumerating the event or subject comprised in the summary and describing the ID and then describes event based summary and permits the users to browse the video summary~~video~~ by the event or subject described in ~~said~~the SummaryThemeList

- 25 It is more preferable that the SummaryThemeList DS ~~includes—include an~~ arbitrary number of SummaryThemes as elements and ~~said~~the SummaryTheme includes an attribute of id representing the corresponding event or subject, and the SummaryTheme further includes an attribute of parentID which is to describe the id of the event or subject of the upper level.

Preferably, the HighlightLevel DS includes an attribute of themelds describing ~~said~~the attribute of ids of common events or subjects if all of the HighlightSegments and HighlightLevels which are constituting the corresponding highlight level have common events or subjects.

- 5 More preferably, the HighlightSegment DS includes an attribute of themelds describing ~~said~~the attribute of id and describes the event or subject of the corresponding highlight segment.

- Also, according to the present invention, a computer-readable recording medium where a HierarchicalSummary DS is stored therein is provided. Preferably, the
- 10 HierarchicalSummary DS comprises at least one HighlightLevel DS, which is describing the highlight level, and the HighlightLevel DS comprises at least one HighlightSegment DS, which is describing highlight segment information constituting the video summary~~videoe~~ of that the highlight level, and the HighlightSegment DS comprises VideoSegmentLocator DS describing time information or video itself of ~~said~~the
- 15 corresponding highlight segment.

- Also, according to the embodiments of the present invention, a method for generating video summary description data according to video summary description scheme by inputting original video is provided. The method includes the following steps: a
- 20 video analyzing step, which is producing video analysis result by inputting the original video and then analyzing the original video; a summary rule defining step, which is defining the summary rule for selecting ~~summary video interval; summary videovideo~~ summary interval; a video summary interval selecting step, which ~~is constituting~~ constitutes video summary~~videoe~~ interval information by selecting the video interval capable of summarizing video contents from the original video by inputting ~~said~~the original
- 25 video analysis result and ~~said~~the summary rule; and a video summary describing step, which is producing video summary description data according to the HierarchicalSummary DS by inputting the video summary~~videoe~~ interval information output by ~~said~~the video summary~~videoe~~ interval selecting step.

Preferably, the video analyzing step comprises a feature extracting step, which is outputting the types of features and video time interval at which those features are detected by inputting the original video and extracting those features, an event detecting step, which is detecting key events included in the original video by inputting saidthe types of features and video time interval at which those features are detected; and an episode detecting step, which is detecting an episode by dividing the original video into a story flow base on the basis of saidthe detected event:

Preferably, the summary rule defining step provides the types of summary events, which are bases in selecting the video summary video interval, after defining them to saidthe video summary describing step.

More preferably, the method further comprises a representative frame extracting step, which is providing the representative frame to saidthe video summary describing step by inputting saidthe video summary video interval information and extracting representative frame.

More preferably, the method further comprises a representative sound extracting step, which is providing the representative sound to saidthe video summary describing step by inputting saidthe video summary video interval information and extracting representative sound.

Also, according to the embodiments of the present invention, a computer-readable recording medium where a program is stored therein is provided. The program executes the following steps: a feature extracting step, which is outputting the types of features and video time interval at which those features are detected; an event detecting step, which is detecting key events included in the original video by inputting saidthe types of features and saidthe video time interval at which those features are detected; an episode detecting step, which is detecting an episode by dividing the original video into a story flow base on the basis of saidthe detected key events; a summary rule defining step, which is defining the summary rule for selecting the summary video interval; summary video video summary interval; a video summary interval selecting step, which is constituting a video summary video interval information by selecting the video interval

capable of summarizing the video contents of the original video by inputting saidthe detected episode and saidthe summary rule; and a video summary describing step, which is generating video summary description data with HierarchicalSummary DS by inputting the video summaryvideo interval information output by saidthe video summaryvideo interval selecting step.

Also, according to the present invention, a system for generating video summary description data according to video summary description scheme by inputting original video is provided. The system includes video analyzing means for outputting a video analysis result by inputting original video and analyzing the original video, summary rule defining means for defining the summary rule for selecting the summary-video-interval, summary-videovideo summary interval, video summary interval selecting means for constituting video summaryvideo interval information by selecting the video interval capable of summarizing the video contents of the original video by inputting saidthe video analysis result and saidthe summary rule, and video summary describing means for generating video summary description data with HierarchicalSummary DS by inputting the video summaryvideo interval information output by saidthe video summaryvideo interval selecting means.

Preferably, the HierarchicalSummary DS comprises at least one HighlightLevel DS, which is describing highlight level, the HighlightLevel DS comprises at least one HighlightSegment DS, which is describing highlight segment information constituting the video summaryvideo of the highlight level, and the HighlightSegment DS comprises VideoSegmentLocator DS describing time information or the video itself of saidthe corresponding highlight segment.

Preferably, the video analyzing means comprises feature extracting means for outputting the types of features and video time interval at which those features are detected by inputting the original video and extracting those features, event detecting means for detecting key events included in the original video by inputting saidthe types of features and video time interval at which those features are detected; and episode detecting

means for detecting episode by dividing the original video into story flow base on the basis of ~~saidthe~~ detected event.

More preferably, the summary rule defining means provides the types of summary events, which are bases in selecting the video summary~~video~~ interval, after
5 defining them to ~~saidthe~~ video summary describing means.

It is preferable that the system further ~~comprises—comprise~~ representative frame extracting means for providing the representative frame to ~~saidthe~~ video summary describing means by inputting ~~saidthe video~~ summary~~video~~ interval information and extracting representative frame.

10 It is more preferable that the system further ~~comprises—comprise~~ representative sound extracting means for providing the representative sound to ~~saidthe~~ video summary describing means by inputting ~~saidthe video~~ summary~~video~~ interval information and extracting representative sound.

Also, according to the embodiments of the present invention, a computer-
15 readable recording medium where a program is stored therein is provided. The program is for functioning feature extracting means for outputting the types of features and video time interval at which those features are detected, event detecting means for detecting key events included in the original video by inputting ~~saidthe~~ types of features and ~~saidthe~~ video time interval at which those features are detected, episode detecting means for
20 detecting episode by dividing the original video into story flow base on the basis of ~~saidthe~~ detected key events, summary rule defining means for defining the summary rule for selecting the ~~summary video interval, summary video~~video summary interval, video summary interval selecting means for constituting video summary~~video~~ interval information by selecting the video interval capable of summarizing the video contents of
25 the original video by inputting ~~saidthe~~ detected episode and ~~saidthe~~ summary rule, and video summary describing means for generating video summary description data with HierarchicalSummary DS by inputting the video summary~~video~~ interval information output by ~~saidthe video~~ summary~~video~~ interval selecting step.

Also, a Video browsing system in a server/client circumstance according to the present invention is provided. The system includes a server ~~which that~~ is equipped with video summary description data generation system which generates video summary description data on the basis of HierarchicalSummary DS by inputting original video and links ~~said the~~ original video and video summary description data, and a client ~~which that~~ is browsing and navigating video by overview of ~~said the~~ original video and access to the original video of ~~said the~~ server using ~~said the~~ video summary description data.

BRIEF DESCRIPTION OF THE DRAWINGS

The embodiments of the present invention will be explained with reference to the accompanying drawings, in which:

FIG. 1 is a block diagram illustrating a system for generating video summary description data according to the description scheme of the present invention.

FIG. 2 is a drawing that illustrates the data structure of the HierarchicalSummary DS describing the video summary description scheme according to the present invention in UML (Unified Modeling Language).

FIG. 3 is a compositional drawing of a user interface of the tool for playing and browsing of the video summaryvideo inputting the video summary description data described by the same description scheme as FIG. 2.

FIG. 4 is a compositional drawing for the flow of the data and control for hierarchical browsing using the video summaryvideo of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

The present invention will be described in detail by way of a preferred embodiment with reference to accompanying drawings, in which like reference numerals are used to identify the same or similar parts.

FIG. 1 is a block diagram illustrating a system for generating video summary description data according to the description scheme of the present invention.

As illustrated in FIG. 1, the apparatus for generating video description data according to an embodiment of the present invention is composed of a feature extracting part 101, an event detecting part 102, an episode detecting part 103, a video summary video interval selecting part 104, a summary rule defining part 105, a representative frame
5 extracting part 106, a representative sound extracting part 107 and a video summary describing part 108.

The feature extracting part 101 extracts necessary features to generate video summary video by inputting the original video. The general features include shot boundary, camera motion, caption region, face region and so on.

10 In the step of extracting features, the types of features and video time interval at which those features are detected are output to the step of detecting event in the format of (types of features, feature serial number, time interval) by extracting those features.

For example, in the case of camera motion, (camera zoom, 1, 100 ~ 150)
15 represents the information that the first zoom of camera was detected in the 100 ~ 150 frame.

The event detecting part 102 detects key events ~~which-that~~ are included in the original video. Because these events must represent the contents of the original video well and are the references for generating ~~summary-video, video summary, these-These~~
20 events are generally differently defined according to genre of the original video.

These events either may represent higher meaning level or may be visual features ~~which-that~~ can directly infer higher meaning. For example, in the case of soccer video, goal, shoot, caption, replay and so on can be defined as events.

The event detecting part 102 outputs the types of detected events and the
25 time interval in the format of (types of events, event serial number, time interval). For example, the event information indicating that the first goal occurred at between 200 and 300 frame is output in the format of (goal, 1, 200 ~ 300).

The episode detecting part 103, on the basis of the detected event, divides the video into an episode with a larger unit than an event based on the story flow. After

detecting key events, an episode is detected while including accompanied events ~~which~~that follow the key event. For example, in the case of soccer video, the goal and shoot can be key events and the bench scene, audiences scene, goal ceremony scene, replay of goal scene and so on compose accompanied events of the key events.

5 That is, the episode is detected on the basis of the goal and shoot.

 The episode detection information is output in the format of (episode number, time interval, priority, feature shot, associated event information). Herein, the episode number is a serial number of the episode and the time interval represents the time interval of the episode by the shot unit. The priority represents the degree of importance of
10 the episode. The feature shot represents the shot number including the most important information out of the shots comprising the episode and the associated event information represents the event number of the event related to the episode. For example, in the case of representing the episode detection information as (episode 1, 4 ~ 6, 1, 5, goal 1, caption 3), the information means that the first episode includes 4 ~ 6th shot, the priority is the highest
15 (1), the feature shot is the fifth shot, and the associated events are the first goal and the third caption.

 The video ~~summary~~video interval selecting part 104 selects the video interval at which the contents of the original video can be summarized well on the basis of the detected episode. The reference of selecting the interval is performed by the predefined
20 summary rule of the summary rule defining part 105.

 The summary rule defining part 105 defines rule for selecting the summary interval and outputs control signal for selecting the summary interval. The summary rule defining part 105 also outputs the types of summary events, which are bases in selecting the video ~~summary~~video interval, to the video summary describing part 108.

25 The video ~~summary~~video interval selecting part 104 outputs the time information of the selected video ~~summary~~video intervals by frame units and outputs the types of events corresponding to the video intervals. That is, the format of (100 ~ 200, goal), (500 ~ 700. shoot) and so on represent that the video segments selected as the video ~~summary~~video intervals are 100 ~ 200 frame, 500 ~ 700 frame and so on and the event of

each segment is goal and shoot respectively. As well, the information such as file name can be output to facilitate the access of an additional video, which is composed of only the video_summaryvideo interval.

5 If the video_summaryvideo interval selection is completed, the representative frame and the representative sound are extracted from the representative frame extracting part 106 and the representative sound extracting part 107 respectively by using the video_summaryvideo interval information.

The representative frame extracting part 106 outputs the image frame number representing the video_summaryvideo interval or outputs the image data.

10 The representative sound extracting part 107 outputs the sound data representing the video_summaryvideo interval or outputs the sound time interval.

The video summary describing part 108 describes the related information in order to make efficient summary and browsing functionalities to be feasible according to the Hierarchical Summary Description Scheme of the present invention shown in FIG. 2.

15 The main information of the Hierarchical Summary Description Scheme comprises the types of summary events of the summary-video, video summary, the time information describing each video_summaryvideo interval, the representative frame, the representative sound, and the event types in each interval.

20 The video summary describing part 108 outputs the video summary description data according to the description scheme illustrated in FIG. 2.

FIG. 2 is a drawing that illustrates the data structure of the HierarchicalSummary DS describing the video summary description scheme according to the present invention in UML (Unified Modeling Language).

25 The HierarchicalSummary DS 201 describing the video summary is composed of one or more HighlightLevel DS 202 and one or zero SummaryThemeList DS 203.

The SummaryThemeList DS provides the functionality of the event based summary and browsing by enumeratively describing the information of subject or event constituting the summary. The HighlightLevel DS 202 is composed of the

HighlightSegment DSs 204 as many as the number of the video intervals constituting the video_summary video of that level and zero or several number of HighlightLevel DS.

The HighlightSegment DS describes the information corresponding to the interval of each summary_video video summary. The HighlightSegment DS is composed of
5 one VideoSegmentLocator DS 205, zero or several ImageLocator DSs 206, zero or several SoundLocator DSs 207 and AudioSegmentLocator 208.

~~The followings—following~~ give more detailed description about the HierarchicalSummary DS.

The HierarchicalSummary DS has an attribute of SummaryComponentList,
10 which obviously represents the summary type; and which is comprised ~~by~~ of the HierarchicalSummary DS.

The SummaryComponentList is derived on the basis of the SummaryComponentType and describes by enumerating all comprised SummaryComponentTypes.

15 In the SummaryComponentList, there are five types, such as keyFrames, keyVideoClips, keyAudioClips, keyEvents, and unconstraint.

The keyFrames represents the key frame summary composed of representative frames. The keyVideoClips represents the key video clip summary composed of key video intervals' sets. The keyEvents represents the summary composed of
20 the video interval corresponding to either the event or the subject. The keyAudioClips represents the key audio clip summary composed of representative audio intervals' sets. And, the unconstraint represents the types of summary defined by users except for ~~said~~ the summaries.

Also, in order to describe ~~the event-based—event-based~~ summary, the
25 HierarchicalSummary DS might comprise the SummaryThemeList DS which is enumerating the event (or subject) comprised in the summary and describing the ID.

The SummaryThemeList has arbitrary number of SummaryThemes as elements. The SummaryTheme has an attribute of id of ID type and selectively has an attribute of parentId.

The SummaryThemeList DS permits the users browsing the video summary~~video~~ from the viewpoint of each event or several subjects described in the SummaryThemeList. That is, the application tool inputting description data makes the ~~users to user~~ select the desired subject by parsing the SummaryThemeList DS and
5 providing the information to the ~~users~~user.

At this time, in the case of enumerating these subjects into simple format, if the number of the subjects ~~are is~~ large, it might not be easy to find out the subject desired by the users.

Accordingly, by representing the subject as a tree structure similar to ToC
10 (Table of Content), the users efficiently can do browsing at each subject after finding out the desired subject.

In order to do so, the embodiments of the present invention ~~permits-permit~~ the attribute of parentId being selectively used in the SummaryTheme. The parentId means the upper element (upper subject) in the tree structure.

15 The HierarchicalSummary DS of the present invention comprises HighlightLevel DSs, and each HighlightLevel DS comprises one or more HighlightSegment DS, which corresponds to a video segment (or interval) constituting the summary-video-video summary.

The HighlightLevel DS has an attribute of themeIds of IDREFS type.

20 The themeIds describes the subject and event id, common to the children HighlightLevel DS of corresponding HighlightLevel DS or all HighlightSegment DSs comprised in the HighlightLevel, and the id is described in ~~said~~the SummaryThemeList DS.

The themeIds can denote several events and, when doing event based summary, solve the problem that same id is unnecessarily repeated in all segments
25 constituting the level by having the themeIds representing common subject type in the HighlightSegment constituting the level.

The HighlightSegment DS comprises one VideoSegmentLocator DS and one or more ImageLocator DS, zero or one SoundLocator DS and zero or one AudioSegmentLocator DS.

Herein, the VideoSegmentLocator DS describes the time information or video itself of the video segment constituting the ~~summary-video:video summary~~. The ImageLocator DS describes the image data information of the representative frame of the video segment. The SoundLocator DS describes the sound information representing the
5 corresponding video segment interval. The AudioSegmentLocator DS describes the interval time information of the audio segment constituting the audio summary or the audio information itself.

The HighlightSegment DS has an attribute of themeIds. The themeIds describes using the id defined in the SummaryThemeList which subjects or events
10 described in said the SummaryThemeList DS relates to the corresponding highlight segment.

The themeIds can denote more than one ~~events-event~~, and by allowing one highlight segment to have several subjects, it is an efficient technique of the present invention which is solving the problem of indispensable duplication of descriptions caused by describing the video segment at each event (or subject) when using the existing method
15 for ~~event-based-event-based~~ summary.

When describing the highlight segment constituting the ~~summary video:video summary~~, in a different way from the existing hierarchical summary description scheme describing only the time information of the highlight video interval, in order to describe the video interval information of each highlight segment, the
20 representative frame information and the representative sound information, by placing the VideoSegmentLocator DS, the ImageSegmentLocator DS and the SoundLocator DS, the present invention makes the overview through the highlight segment video and the navigation and browsing utilizing the representative frame and the representative sound of the segment to be feasible to efficiently utilize through the introduction of the
25 HighlightSegment DS for describing the highlight segment constituting the ~~summary video:video summary~~.

By placing the SoundLocator DS capable of describing the representative sound corresponding to the video interval, in real instances through the characteristic sound capable of representing the video interval, for example gun shot, outcry, anchor's comment

in soccer (for example, goal and shoot), actors' name in drama, specific word, etc., it is possible to do efficient browsing by roughly understanding whether the interval is an important interval containing the desired contents or what contents are contained in the interval within a short time without playing the video interval.

5 FIG. 3 is a compositional drawing of a user interface of the tool for playing and browsing of the video summary~~video~~ inputting the video summary description data described by the same description scheme as FIG. 2.

 The video playing part 301 plays the original video or the video summary~~video~~ according to the control of the user. The original video representative frame
10 part 305 shows the representative frames of the original video shots. That is, it is composed of a series of images with reduced sizes.

 The representative frame of the original video shot is described not by the HierarchicalSummary DS of the present invention but by additional description scheme and can be utilized when both the description data are provided along with the summary
15 description data described by the HierarchicalSummary DS of the present invention.

 The user accesses to the original video shot corresponding to the representative frame by clicking the representative frame.

 The video summary~~video~~ level 0 representative frame part and the representative sound part 307 and the video summary~~video~~ level 1 representative frame
20 part and the representative sound part 306 shows the frame and sound information representing each video interval of the video summary~~video~~ level 0 and the video summary
~~video~~-level 1 respectively. That is, it is composed of the iconic images representing a series of the images and sounds with reduced sizes.

 If the user clicks the representative frame of the video summary~~video~~
25 representative frame part and the representative sound part, the user accesses to the original video interval corresponding to the representative frame. Herein, in the case of clicking the representative sound icon corresponding to the representative frame of the summary
video,video summary, the representative sound of the video interval is played.

The video summaryvideo controlling part 302 inputs the control for user selection to play the summary-video:video summary. In the case of being provided with the multi-level-multi-level summary-video:video summary, the user does overview and browsing by selecting the summary of the desired level through the level selecting part 303.

- 5 The event selecting part 304 enumerates the event and the subject provided by the SummaryThemeList and the user does overview and browsing by selecting the desired event. After all, this realizes the summary of the user customization type.

FIG. 4 is a compositional drawing for the flow of the data and control for hierarchical browsing using the video summaryvideo of the present invention.

- 10 The browsing is performed by accessing the data for browsing with the method of FIG. 4 through the use of the user interface of FIG. 3. The data for browsing are the video summaryvideo and the representative frame of the video summaryvideo and the original video 406 and the original video representative frame 405.

- The video summaryvideo is assumed to have two levels. Needless to say, 15 the video summaryvideo may have more levels than two. The video summaryvideo level 0 401 is what is summarized with shorter time than the video summaryvideo level 1 403. That is, the video summaryvideo level 1 contains more contents than the summary-video:video summary level 0. The summary-video:video summary level 0 representative frame 402 is the representative frame of the video summaryvideo level 0 and the video 20 summaryvideo level 1 representative frame 404 is the representative frame of the summary-video:video summary level 1.

- The summary-video:video summary and the original video are played through the video playing part 301 ~~of shown in~~ FIG. 3. The video summaryvideo level 0 representative frame is displayed in the video summaryvideo level 0 representative frame 25 and the representative sound part 306, the video summaryvideo level 1 representative frame is displayed in the video summaryvideo level 1 representative frame and the representative sound part 307, and the original video representative frame is displayed in the original video representative frame part 305.

The hierarchical browsing method illustrated in FIG. 4 can have various types of hierarchical paths as the following example.

Case 1 : (1) - (2)

Case 2 : (1) - (3) - (5)

5 Case 3 : (1) - (3) - (4) - (6)

Case 4 : (7) - (5)

Case 5 : (7) - (4) - (6)

The overall browsing scheme is as follows.

10 First, understand the overall contents of the original video by watching the video_summaryvideo of the original video. Herein, the video_summaryvideo may play either the video_summaryvideo level 0 or the video_summaryvideo level 1. When more detailed browsing is wanted after watching the summary-video, video_summary, the interested video interval is identified through the video_summaryvideo representative frame. If the scene which is desired to be exactly found, is identified in the video_summaryvideo representative frame, play it by directly accessing to the video interval of the original video to which the representative frame is connected. And if the more detailed information is needed, the user may access to the desired original video either by understanding the representative frame of the next level or by hierarchically understanding the contents of the representative frame of the original video.

20 Although these hierarchical browsing techniques might take a long time in browsing to access to the desired contents while the original video is being played, the browsing time is ~~drastically~~ substantially reduced by directly accessing to the contents of the original video through the hierarchical representative frame.

25 The existing general video indexing and browsing techniques divide the original video in shot unit and access to the shot by perceiving the desired shot from the representative frame after constituting the representative frame representing each shot.

In this case, because the number of the shots of in the original video is large, ~~lots of substantial~~ time and efforts are necessary to do browsing the desired contents out of many representative frames.

In the present invention, it is feasible to quickly access to the desired video by constituting the hierarchical representative frame with the representative frame of the summary video-video summary.

5 The case 1 is the case that plays the video summary level 0 and directly accesses to the original video from the video summary level 0 representative frame.

10 The case 2 is the case that plays the video summary level 0 and selects the most interested representative frame from the video summary level 0 representative frame and identifies the desired scene in the video summary level 1 representative frame corresponding to the neighborhood of the representative frame to understand more detailed information before access to the original video and then accesses to the original video.

15 The case 3 is the case that selects the most interested representative frame to obtain more detailed information in the case that the access from the video summary level 1 representative frame to the original video is difficult in the case 2 and by the original video representative frames neighboring the representative frame identifies the desired scene and then accesses to the original video using the representative frame of the original frame.

20 The case 4 and case 5 are the cases that start at the playing of the video summary level 1 and the paths are similar to the above cases.

25 When applied to the server/client circumstance, the present invention can provide ~~the a~~ system in which multiple clients can access to one server and ~~can~~ do video overview and browsing. The original video is inputted to the server and the video summary description data is produced on the basis of the hierarchical summary description scheme and the video summary description data generation system linking ~~said~~ the original video and the video summary description data is equipped. The client accesses to the server through the communication network, does overview of the video using the video summary description data, and does browsing and navigation of the video by accessing to the original video.

Although, the present invention was described on the basis of preferably executable examples, these executable examples do not limit the present invention but exemplify. Also, it will be appreciated by those skilled in the art that changes and variations in the embodiments herein can be made without departing from the spirit and
5 scope of the present invention as defined by the following claims and the equivalents thereof.

300055.468 / 530420_1.DOC

Amendments to the Claims:

This listing of claims will replace all prior versions, and listings, of claims in the application:

Listing of Claims:

1. (Currently Amended) A HierarchicalSummary Description Scheme (DS) for describing a video summary, the HierarchicalSummary DS ~~comprises comprising:~~ at least one HighlightLevel DS which is describing highlight level, wherein said a HighlightLevel DS that comprises at least one HighlightSegment DS, which is describing the HighlightSegment DS configured to describe highlight segment information constituting the summary video of the on a highlight level segment corresponding to one of a plurality of video summary intervals, the HighlightSegment DS further comprising a VideoSegmentLocator DS describing time information or a video itself of the highlight segment and an ImageLocator DS describing a representative frame of the highlight segment.
2. (Currently Amended) The HierarchicalSummary DS ~~according to of~~ claim 1; wherein said HighlightLevel DS ~~is composed of further comprises~~ at least one lower level HighlightLevel DSs DS.
- 3.-4. (Canceled)
5. (Currently Amended) The HierarchicalSummary DS ~~according to of~~ claim 3; wherein said HighlightSegment DS further comprises a SoundLocator DS which is describing that describes the representative sound information of said corresponding highlight segment.
- 6.-7. (Canceled).

8. (Currently Amended) The HierarchicalSummary DS ~~according to~~ of claim 3; wherein said HighlightSegment DS further comprises an AudioSegmentLocator DS which is describing that describes the audio segment information constituting an audio summary of said corresponding highlight segment.

9. (Canceled)

10. (Currently Amended) The HierarchicalSummary DS ~~according to~~ of claim ~~8~~⁹; wherein ~~said the~~ HierarchicalSummary DS includes an attribute of a SummaryComponentList describing and enumerating all of the that describes and enumerates SummaryComponentTypes which is included in the HierarchicalSummary DS representing types of summary.

11. (Currently Amended) The HierarchicalSummary DS ~~according to~~ of claim 10; wherein said SummaryComponentType ~~includes~~ comprises keyFrames representing ~~the a~~ key frame summary composed of representative frames, keyVideoClips representing ~~the a~~ key video clip summary composed of key video segment' sets, keyEvents representing ~~the a~~ summary of ~~the a~~ video interval corresponding to either ~~the an~~ event or ~~the a~~ subject, and keyAudioClips representing ~~the a~~ key audio clip summary composed of representative audio intervals' sets, and an unconstraint representing the type of summary defined by users except for said summaries.

12. (Currently Amended) The HierarchicalSummary DS ~~according to~~ of claim 1; wherein ~~said the~~ HierarchicalSummary DS further includes a SummaryThemeList DS which is enumerating the event or subject comprised in the summary and describing the ID and then describes event based summary and permits the enabling users to browse the summary video by the event or subject described in said SummaryThemeList execute summarizing and browsing based on the event

13. (Currently Amended) The HierarchicalSummary DS ~~according to of~~ claim ~~12+1~~, wherein ~~said the~~ SummaryThemeList DS ~~includes comprises an~~ arbitrary number of SummaryThemes as elements, and ~~said the~~ SummaryTheme includes an attribute of and id-ID representing ~~the a~~ corresponding event or subject.

14. (Currently Amended) The HierarchicalSummary DS ~~according to of~~ claim 13; wherein ~~said the~~ SummaryTheme further includes an attribute of parentID ~~which is to describe that describes~~ the id-ID of the event or subject of the upper level.

15. (Currently Amended) The HierarchicalSummary DS ~~according to of~~ claim 13; wherein ~~said the~~ HighlightLevel DS ~~includes comprises an~~ attribute of themelds themeIDs describing ~~said an~~ attribute of ids of common events or subjects if all of the ID, when the HighlightSegments ~~and or~~ HighlightLevels ~~which are constituting that constitute a~~ corresponding highlight level have common events or subjects, the ID of the common events or the subjects is described in the themeID.

16. (Currently Amended) The HierarchicalSummary DS ~~according to of~~ claim 13; wherein ~~said the~~ HighlightSegment DS ~~includes comprises an~~ attribute of themelds themeIDs describing ~~said attribute of id and describes the event or subject of the corresponding~~ highlight segment using the attribute of ID described in the SummaryThemeList DS.

17. (Currently Amended) A computer-readable recording medium where a HierarchicalSummary DS for describing a video summary is stored therein, wherein the HierarchicalSummary DS comprises at least one a HighlightLevel DS which is describing that highlight level, wherein ~~said HighlightLevel DS comprises at least one HighlightSegment DS,~~ the HighlightSegment DS describes information on a ~~which is describing highlight segment~~ corresponding to one of information constituting the a plurality of summary-video summary ~~intervals, and the of that the highlight level, wherein said HighlightSegment DS further~~ comprises a VideoSegmentLocator DS describing time information or a video itself of said

corresponding the highlight segment and an ImageLocator DS describing a representative frame of the highlight segment.

18. (Currently Amended) A method for generating video summary description data according to a video summary description scheme by inputting original video, comprising:

(a) video analyzing step which is producing video analysis result by inputting analyzing the input the original video and then analyzing the original video producing video analysis result;

(b) summary rule defining step which is defining the a summary rule for selecting the video summary video interval;

(c) summary video interval selecting step which is constituting summary video interval information by selecting the video summary interval capable of summarizing video contents from the original video by inputting said based on the original video analysis result and said the summary rule and constituting video summary interval information; and

(d) extracting a representative frame based on the video summary interval information; and

(e) video summary describing step which is producing generating video summary description data according to the HierarchicalSummary DS enabling to execute hierarchical browsing based on by inputting the video summary video interval information output by said summary video interval selecting step and the representative frame.

wherein the HeirarchicalSummary DS comprises a HighlightLevel DS that comprises at least one HighlightSegment DS, and the HighlightSegment DS describes information on a highlight segment corresponding to one of a plurality of video summary intervals, and the HighlightSegment DS comprises a VideoSegmentLocator DS describing time information or a video itself of the highlight segment.

19. (Canceled)

20. (Currently Amended) The method of claim 18 wherein for generating video summary description data according to claim 18, wherein said video analyzing step (a) comprises the steps of:

feature extracting step which is outputting the types of extracting features and video time interval at which those features are detected by inputting the from the input original video and extracting those outputting the types of features and video time interval at which those features are detected;

event detecting step which is detecting key events included in the original video by inputting said based on the types of features and video time interval at which those features are detected; and

episode detecting step which is detecting episode by dividing the original video into story flow base on the basis of said the detected event.

21.-22. (Canceled)

23. (Currently Amended) The method for ~~generating video summary description data according to of~~ claim 18; wherein the method further comprises representative sound extracting step (d) which is providing the representative sound to said video summary describing comprises the step by inputting said summary video interval information and of extracting a representative sound from the video summary interval information.

24. (Canceled)

25. (Currently Amended) A system for generating video summary description data according to a video summary description scheme by inputting original video, comprising:

video analyzing means for outputting video analysis result by inputting original video and analyzing the original video and producing video analysis result;

summary rule defining means for defining the summary rule for selecting the summary-video summary interval;

~~video summary video~~-interval selecting means for ~~constituting summary video interval information by~~ selecting the video interval capable of summarizing the video contents of the original video and outputting video summary interval information based on the by-inputting said-video analysis result from the video analyzing means and said-the summary rule from the summary rule defining means; and

representative frame extracting means for outputting a representative frame representing video summary interval based on the video summary interval information from the video summary interval selecting means; and

video summary describing means for generating video summary description data with HierarchicalSummary DS by inputting the ~~summary-video summary~~ interval information from the output by said summary-video summary interval selecting means and the representative frame information from the representative frame extracting means.

wherein the HierarchicalSummary DS comprises a HighlightLevel DS that comprises at least one HighlightSegment DS, and the HighlightSegment DS describes information on segment information corresponding to one of a plurality of video summary intervals, and the HighlightSegment DS comprises a VideoSegmentLocator DS describing time information or a video itself of the highlight segment and an ImageLocator DS describing a representative frame of the highlight segment.

26. (Canceled)

27. (Currently Amended) The system for ~~generating video summary description data according to of~~ claim 25; wherein said video analyzing means comprises:

feature extracting means for outputting the types of features and video time interval at which those features are detected by inputting the extracting features from the original video and extracting those-producing the types of features and video time interval at which those features are detected;

event detecting means for detecting key events included in the original video by inputting ~~said the~~ types of features and video time interval at which those features are detected; and

episode detecting means for detecting episode by dividing the original video into story flow base on the basis of said detected event.

28.-29. (Canceled)

30. (Currently Amended) The system ~~for generating video summary description data according to of~~ claim 25, the system further comprises representative sound extracting means for ~~providing the extracting a~~ representative sound to ~~said video summary describing means by~~ inputting ~~said the summary video summary interval information and extracting providing the extracted~~ representative sound to the video summary describing means.

31.-32. (Canceled)

33. (New) The method of claim 18 wherein the HighlightSegment DS further comprises an ImageLocator DS describing a representative frame of the highlight segment.

34. (New) The method of claim 18 wherein the HighlightSegment DS further comprises SoundLocator DS describing a representative sound information of the highlight segment.

35. (New) The method of claim 18 wherein the HighlightSegment DS further comprises AudioSegmentLocator DS describing the audio segment information constituting an audio summary of the highlight segment.

36. (New) The method of claim 18 wherein the HierarchicalSummary DS further includes SummaryThemeList DS enumerating the event or subject comprised in the summary and enabling user's to execute summarizing and browsing based on the event.
37. (New) The method of claim 36 wherein the SummaryThemeList DS includes arbitrary number of SummaryThemes as elements, and the SummaryTheme includes an attribute of ID representing the corresponding event or subject.
38. (New) The method of claim 37 wherein the SummaryTheme further includes an attribute of parentID describing the ID of the event or subject of the upper level.
39. (New) The method of 37 wherein the HighlightLevel DS includes an attribute of themeIDs describing attribute of the ID, when the HighlightSegments or HighlightLevels which are constituting corresponding highlight level have common events or subjects, the ID of the common events or the subjects is described in the themeID.
40. (New) The method of claim 37 wherein the HighlightSegment DS includes an attribute of themeIDs describing the event or subject of the highlight segment using the attribute of ID described in the SummaryThemeList DS.
41. (New) The system of claim 25 wherein the HighlightSegment DS further comprises SoundLocator DS describing a representative sound information of the highlight segment.
42. (New) The system of claim 25 wherein the HighlightSegment DS further comprises AudioSegmentLocator DS describing the audio segment information constituting an audio summary of the highlight segment.

43. (New) The system of claim 25 wherein the HierarchicalSummary DS further includes SummaryThemeList DS enumerating the event or subject comprised in the summary and enabling user's to execute summarizing and browsing based on the event.

44. (New) An apparatus for browsing video summary description data, wherein the video summary description data has a HierarchicalSummary Description Scheme (DS) for describing a video summary, the HierarchicalSummary DS includes: a HighlightLevel DS which comprises at least one HighlightSegment DS describing information on highlight segment corresponding one of video summary intervals, and a SummaryThemeList DS enumerating the event or subject comprised in the summary and enabling user's to execute summarizing and browsing based on the event, wherein the HighlightSegment DS includes a VideoSegmentLocator DS describing time information or video itself of the highlight segment, and an ImageLocator DS describing a representative frame of the highlight segment, wherein the browsing apparatus comprising:

- a video playing part for playing an original video or the video summary;

- an original video representative frame part for playing a representative frame of the original video;

- a first video summary representative frame part for playing a first summary level of video interval,

- a second video summary representative frame part for playing a second summary level of video interval, wherein the second summary level is summarized more finely than the first summary level;

- a level selecting part for selecting the first summary level or the second summary level thereby enabling the video playing part to play the selected summary level; and

- an event selecting part for enumerating the event or the subject provided in the SummaryThemeList DS for a user to browse desired event.

45. (New) The apparatus of claim 44 wherein the first video summary representative frame part plays the first summary level of sound information, and the second video summary representative frame part plays the first summary level of sound information.

46. (New) The apparatus of claim 45 wherein the HighlightSegment DS further comprises:

a SoundLocator DS describing a representative sound information of the highlight segment; and

an AudioSegmentLocator DS describing the audio segment information constituting an audio summary of the highlight segment.

47. (New) A method of browsing video summary description data, wherein the video summary description data having a HierarchicalSummary Description Scheme (DS) for describing a video summary,

wherein the HierarchicalSummary DS comprises a HighlightLevel DS which comprises at least one HighlightSegment DS describing information on highlight segment corresponding one of video summary intervals, and

wherein the HighlightSegment DS comprises a VideoSegmentLocator DS describing time information or video itself of the highlight segment, and an ImageLocator DS describing a representative frame of the highlight segment,

wherein the browsing method comprises:

- (a) playing a first summary level of video summary; and
- (b) playing video interval of original video corresponding to the representative frame when a desired scene is found through the video summary representative frame at the step (a).

48. (New) The method of claim 47, further comprising:

(c) playing a second summary level of video summary when a desired scene is not found through the video summary representative frame at the step (a), wherein the second summary level is summarized more finely than the first summary level.

49. (New) The method of claim 47 wherein the HighlightSegment DS further comprises a SoundLocator DS describing a representative sound information of the highlight segment, the step (b) comprising the step of recognizing the desired scene to be found at the step (a) through the video summary representative sound information.

50. (New) The method of claim 47 wherein the HierarchicalSummary DS further includes SummaryThemeList DS enumerating the event or subject comprised in the summary and enabling user's to execute summarizing and browsing based on the event.

51. (New) A Video Summary Description Scheme (DS) for describing a video summary, comprising:

at least one HighlightSegment DS describing information on highlight segment corresponding one of video summary intervals, wherein the HighlightSegment DS comprises a VideoSegmentLocator DS describing time information or video itself of the highlight segment and an ImageLocator DS describing a representative frame of the highlight segment.

52. (New) The Video Summary DS of claim 51 wherein the HighlightSegment DS further comprises SoundLocator DS describing a representative sound information of the highlight segment.

53. (New) The Video Summary DS of claim 51 wherein the HighlightSegment DS further comprises AudioSegmentLocator DS describing the audio segment information constituting an audio summary of the highlight segment.

54. (New) A method of browsing video summary description data wherein the video summary description data having at least one HighlightSegment DS describing information on highlight segment corresponding one of video summary intervals, wherein the HighlightSegment DS comprises a VideoSegmentLocator DS describing time information or video itself of the highlight segment and an ImageLocator DS describing a representative frame of the highlight segment,

wherein the browsing method comprising:

- (a) playing a first summary level of video summary; and
- (b) playing video interval of original video corresponding to the representative frame when a desired scene is found through the video summary representative frame at the step (a).

55. (New) The method of claim 54 further comprising:

- (c) playing a second summary level of video summary when a desired scene is not found through the video summary representative frame at the step (a), wherein the second summary level is summarized more finely than the first summary level .

56. (New) The method of claim 54 wherein the HighlightSegment DS further comprises a SoundLocator DS describing a representative sound information of the highlight segment, the step (b) comprising the step of recognizing the desired scene to be found at the step (a) through the video summary representative sound information.

Amendments to the Abstract:

Please replace the previous Abstract with the following redlined Abstract:

**VIDEO SUMMARY DESCRIPTION SCHEME AND METHOD AND SYSTEM OF VIDEO
SUMMARY DESCRIPTION DATA GENERATION FOR EFFICIENT OVERVIEW AND
BROWSING**

ABSTRACT OF THE DISCLOSURE

The present invention relates to a A video summary description scheme system for describing video summary by meta data. The video summary to provides overview functionality; which that makes it feasible to understand the overall contents of the original video within a short time and with navigation and browsing functionalities, which make and that makes it feasible to search the desired video contents efficiently. According to the present invention the The system includes a HierarchicalSummary Description Scheme (DS) comprises at least one for describing a video summary that includes a HighlightLevel DS and selectively comprises the SummaryThemeList DS. The HighlightLevel DS describes highlight level and may have zero or at least one lower HighlightLevel DS. The HighlightLevel DS comprises one or more having at least one HighlightSegment DS which is describing highlight segment information constituting the video summary of the highlight level. The HighlightSegment DS comprises the VideoSegmentLocator DS for describing the time describes information of corresponding on a highlight segment interval. Also, the HighlightSegment DS may comprise the ImageLocator DS for describing the representative image information of corresponding segment, the SoundLocator DS for describing the representative sound information, and the AudioSegmentLocator DS for describing the audio segment information constituting the audio summary corresponding to one of the video summary intervals, and the HighlightSegment DS includes a VideoSegmentLocator DS for describing time information or the video itself of the

highlight segment and an ImageLocator DS describing a representative frame of the highlight segment.